A Comprehensive Exploration on WikiSQL with Table-Aware Word Contextualization

Wonseok Hwang  Jinyeung Yim  Seunghyun Park  Minjoon Seo
Clova AI, NAVER Corp.
{wonseok.hwang,jinyeong.yim,seung.park,minjoon.seo} @navercorp.com

Abstract

WikiSQL is the task of mapping a natural language question to a SQL query given a table from a Wikipedia article. We first show that learning highly context-aware word representations is arguably the most important consideration for achieving a high accuracy in the task. We explore three variants of BERT-based architecture and our best model outperforms the previous state of the art by 8.2% and 2.5% in logical form and execution accuracy, respectively. We provide a detailed analysis of the models to guide how word contextualization can be utilized in such a semantic parsing task. We then argue that this score is near the upper bound in WikiSQL, where we observe that most of the evaluation errors are due to wrong annotations. We also measure human accuracy on a portion of the dataset and show that our model exceeds the human performance, at least by 1.4% execution accuracy.

1 Introduction

Semantic parsing is the task of translating natural language utterances to (often machine-executable) formal meaning representations. By helping non-experts to interact with ever-increasing databases, the task has many important potential applications in real life such as question answering (Guo et al., 2018) and navigation control (Gupta et al., 2018) via speech-based smart devices.

Despite the importance of the task, semantic parsing datasets have suffered from the lack of full (logical form) annotations, which often need many expert-hours to directly obtain them. Zhong et al. (2017) recently introduced WikiSQL as one of the first large-scale semantic parsing datasets, with 80,654 pairs of questions and the corresponding human-verified SQL queries. The massive dataset has attracted much attention in the community and witnessed a significant progress through task-specific end-to-end neural models (Xu et al., 2017).

On the other side of natural language processing community, we have also observed a rapid advancement in contextualized word representations (Peters et al., 2018; Devlin et al., 2018), which proved to be extremely effective for most language tasks that deal with unstructured text data. However, it has not been clear yet whether the word contextualization is also similarly effective when structured data such as tables in WikiSQL are involved.

In this paper, we discuss our approach on WikiSQL with BERT (Devlin et al., 2018) as the backbone and provide a comprehensive analysis of the dataset using our model. In particular, we propose table-aware BERT encoder (Section 3) and three different modules on top of the encoder for the task-specific part (Section 4), in the order of increasing complexity: SHALLOW-LAYER, DECODER-LAYER, and NL2SQL-LAYER. We show that even a minimal module (SHALLOW-LAYER) outperforms the previous best model, but we also see a better and a more robust performance with a dense module (NL2SQL-LAYER), achieving 83.6% logical form accuracy and 89.6% execution accuracy on WikiSQL test set, outperforming the previous best model by 8.2% and 2.5%, respectively. We furthermore argue that these scores are near the
upper bound in WikiSQL, where we observe that most of the evaluation errors are due to wrong annotations by humans. In fact, according to our turked statistics on an approximately 10% sample of WikiSQL dataset, our model’s score exceeds human performance at least by 1.4% in execution accuracy.

Our key contributions are summarized as follows:

- We verify that word contextualization is also crucial for language tasks with structured data. Our proposed BERT-based table-aware encoder and task-specific modules outperform the previous best model in WikiSQL.

- We show that our models effectively achieve the upper bound of the accuracy on WikiSQL task. We back this argument by performing a detailed error analysis and human performance evaluation.

Our source code is publicly available at https://github.com/naver/sqlova.

2 Related Work

WikiSQL is a large semantic parsing dataset consisting of 80,654 natural language utterances and corresponding SQL annotations on 24,241 tables extracted from Wikipedia (Zhong et al., 2017). The task is to build the model that generates SQL query for given natural language question on single table and table headers without using contents of the table. Some examples, using the table from WikiSQL, are shown in Table 1.

Table 1: Example of WikiSQL semantic parsing task. For given questions and table headers, the model generates corresponding SQL query and retrieves the answer from the table.

<table>
<thead>
<tr>
<th>Player</th>
<th>Country</th>
<th>Points</th>
<th>Winnings ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steve Stricker</td>
<td>United States</td>
<td>9000</td>
<td>1260000</td>
</tr>
<tr>
<td>K.J. Choi</td>
<td>South Korea</td>
<td>5400</td>
<td>756000</td>
</tr>
<tr>
<td>Rory Sabbatini</td>
<td>South Africa</td>
<td>3400</td>
<td>4760000</td>
</tr>
<tr>
<td>Mark Calcavecchia</td>
<td>United States</td>
<td>2067</td>
<td>289333</td>
</tr>
<tr>
<td>Ernie Els</td>
<td>South Africa</td>
<td>2067</td>
<td>289333</td>
</tr>
</tbody>
</table>

Question: What is the points of South Korea player?
SQL: SELECT Points WHERE Country = South Korea
Answer: 5400

The large size of the dataset has enabled adopting deep neural techniques for the task and drew much attention in the community recently. Although early studies on neural semantic parsers have started without syntax specific constraints on output space (Dong and Lapata, 2016; Jia and Liang, 2016; Iyer et al., 2017), many state-of-the-arts results on WikiSQL have been achieved by constraining the output space with the SQL syntax. The initial model proposed by (Zhong et al., 2017) independently generates the two components of the target SQL query, select-clause and where-clause which outperforms the vanilla sequence-to-sequence baseline model proposed by the same authors. SQLNet (Xu et al., 2017) further simplifies the generation task by introducing a sequence-to-set model in which only where condition value is generated by the sequence-to-sequence model, hence making the model insensitive to the order of the SQL conditions. TypeSQL (Yu et al., 2018) also employs a sequence-to-set structure but with an additional “type” information of natural language tokens. Coarse2Fine (Dong and Lapata, 2018) first generates rough intermediate output, and then refines the results by decoding full where-clauses. Similarly to our NL2SQL-LAYER, the final table-aware contextual representation of the question is generated with bi-LSTM with attention. Our model however differs in that many layers of self-attentions (Vaswani et al., 2017; Devlin et al., 2018) are employed with a single concatenated input of question and table headers. Pointer-SQL (Wang et al., 2017) proposes a sequence-to-sequence model that uses an attention-based copying mechanism and a value-based loss function. Annotated Seq2seq (Wang et al., 2018b) utilizes a sequence-to-sequence model after automatic annotation of input natural language. MQAN (McCann et al., 2018) suggests a multitask question answering network that jointly learns multiple natural language processing tasks using various attention mechanisms. Execution guided decoding is suggested in ref. (Wang et al., 2018a), in which non-executable (partial) SQL queries candidates are removed from output candidates during decod-
ing step. IncSQL (Shi et al., 2018) proposes a sequence-to-action parsing approach that uses incremental slot filling mechanism with feasible actions from a pre-defined inventory.

### 3 Table-aware BERT Encoder

Although pretrained word representations on a large (unlabeled) language corpus, such as GloVe (Pennington et al., 2014), have shown promising results in WikiSQL (Dong and Lapata, 2018; Xu et al., 2017; Zhong et al., 2017; Shi et al., 2018; Yu et al., 2018; Xiong and Sun, 2018; Wang et al., 2017, 2018b; Yin and Neubig, 2018), recently developed contextualized word representations such as ELMO (Peters et al., 2018) and BERT (Devlin et al., 2018) show superior performances in many NLP tasks. Here, we extend BERT (Devlin et al., 2018) for encoding the natural language query together with the headers of the entire table. We use [SEP] to separate between the query and the headers. That is, each query input $T_{h,1}, \ldots, T_{h,L}$ ($L$ is the number of query words) is encoded as following:

$[CLS], T_{h,1}, \ldots, T_{h,L}, [SEP], T_{h1,1}, T_{h1,2}, \ldots, [SEP], \ldots, [SEP], T_{hN_h,1}, \ldots, T_{hN_h,M_{Nh}}$, [SEP]

**Table-aware BERT-Encoder**

![Table-aware BERT-Encoder](attachment:image.png)

Figure 1: (A) The scheme of input encoding process by table-aware BERT. Final output vectors are represented by colored bars: light blue for [CLS] output, red for question words, and green for tokens from table headers.

where $T_{h,j,k}$ is the $k$-th token of the $j$-th table header, $M_j$ is the total number of tokens of the $j$-th table headers, and $N_h$ is the total number of table headers. Each token consists of token embedding, segment embedding, and position embedding. [CLS] and [SEP] are special tokens for classification and context separation, same as (Devlin et al., 2018).

This input scheme is used in SHALLOW-LAYER and NL2SQL-LAYER. For DECODER-LAYER, additional SQL vocabulary such as select, where, min, and > and start and end tokens are placed between [CLS] and question words separated by [SEP] for the sequence generation. By placing them in front of question- and header-tokens, their positions remain invariant to questions and tables headers. The output from final Transformer block (Vaswani et al., 2017) are used in SHALLOW-LAYER and DECODER-LAYER whereas the output of final two Transformer blocks are concatenated and used in NL2SQL-LAYER.

### 4 Model

In this section, we describe the details of three proposed modules on top of the table-aware BERT encoder: SHALLOW-LAYER, DECODER-LAYER, and NL2SQL-LAYER.

#### 4.1 SHALLOW-LAYER

SHALLOW-Layer does not contain trainable parameters but controls the flow of information during fine-tuning of BERT via loss function. Compared to other types of encoders, SHALLOW-LAYER has a merit of simplicity in use.

In a typical sequence generation model, the output is not explicitly constrained by any syntax, which is highly suboptimal for formal language generation. Hence, following (Xu et al., 2017), SHALLOW-LAYER uses syntax-guided sketch, where the generation model consists of six modules, namely select-column, select-aggregation, where-number, where-column, where-operator, and where-value (Fig. 2). Before describing what each part is responsible for, we first introduce our notations: $H_{[CLS]}$ stands for the output of [CLS] token from table-aware BERT encoder, $H_{n,i}$ for the output of $T_{n,i}$, and $H_{h,i}$ for the output of $T_{h,i}$. All three real vectors belong to $\mathbb{R}^d$ where $d$ is the hidden dimension of the BERT encoder (for example, $d = 1024$ for BERT-Large model). $(H)_\mu$ denotes
Figure 2: (A) The model scheme of SHALLOW-LAYER. Each circle represents the single element of the output vector from table-aware BERT-encoder. The role of each elements in SQL query generation is indicated by black squares. For example, the probability of the word “the” to be the start token of where-value of 1st header is calculated by using 1st element of $H_{\text{the}}$ vector together with 1st elements of all $H$ vectors of question words. (B) The scheme DECODER-LAYER. LSTM-decoder of pointer network (Vinyals et al., 2015) generates the sequence of pointers to augmented inputs which include SQL vocabulary, start, end, question words, and header tokens. Generated pointer sequences are interpreted by Pointer-to-SQL module which generates final SQL queries. (C) The scheme of the information flow in NL2SQL-LAYER (SQLOVA). The outputs from table-aware BERT encoder are encoded again by LSTM-q (question encoder) and LSTM-h (header encoder). In each module, column attention (Xu et al., 2017) is employed.

$\mu$-th element of the vector $H$. $W$ stands for affine transformation. To make the equation uncluttered, same $W$ is used to denote any affine transformation. Also, we represent the conditional probability for given question and table-schema $p(\cdot|Q, \text{table-schema})$ as $p(\cdot)$.

select-column finds the column in select clause from given natural language utterance as follow.

$$s_{sc}(i) = (H_{h,i})_0$$
$$p_{sc}(col_i) = \text{softmax} (s_{sc}(i))$$

$p_{sc}(col_i)$ is the probability of generating $i$-th header in select clause.

select-aggregation finds the aggregation operator for the given select column. For example, if $i$-th header is generated in select clause,

$$s_{sa}(\text{NONE}|col_i) = (H_{h,i})_1$$
$$s_{sa}(\text{MAX}|col_i) = (H_{h,i})_2$$
$$s_{sa}(\text{MIN}|col_i) = (H_{h,i})_3$$
$$s_{sa}(\text{COUNT}|col_i) = (H_{h,i})_4$$
$$s_{sa}(\text{SUM}|col_i) = (H_{h,i})_5$$
$$s_{sa}(\text{AVG}|col_i) = (H_{h,i})_6$$

The probability of generating aggregation operator is calculated by feeding $s_{sa}$ to softmax function.

where-number predicts the number of where conditions in SQL queries.

$$s_{wn}(\mu) = (WH_{\text{CLS}})_\mu$$

The probability of generating $\mu$ number of where conditions are calculated via softmax function.

where-column calculates the probability of generating each columns for given natural
operators for given

\[
s_{\text{wc}}(\text{col}_i) = (H_{h,i})^t
p_{\text{wc}}(\text{col}_i) = \text{sigmoid}(s_{\text{wc}}(\text{col}_i))
\] (4)

\(p_{\text{wc}}\) stands for the probability that \(\text{col}_i\) is generated in where clause.

where-operator finds most probable operators for given where column among three possible choices (\(>, =, <\)). For example for \(i\)-th column,

\[
s_{\text{wo}}(\text{col}_i) = (H_{h,i})_8
s_{\text{wo}}(\text{col}_i) = (H_{h,i})_9
s_{\text{wo}}(\text{col}_i) = (H_{h,i})_{10}
\] (5)

The probability of generating each operator for given column \(i\) is calculated by feeding \(s_{\text{wo}}(\text{col}_i)\) to softmax function.

where-value finds which tokens of a natural language utterance correspond to condition values for given where columns. The probability of \(k\)-th token of the question is selected as start index of where value for given \(\mu\)-th column (\(\text{col}_\mu\)) is calculated as follow.

\[
s_{\text{wv,ed}}(k|\text{col}_\mu) = (H_{n_k})_{\mu+100}
p_{\text{wv,ed}}(k|\text{col}_\mu) = \text{softmax}(s_{\text{wv,ed}}(k|\text{col}_\mu))
\] (7)

100 is selected as the maximum number of headers of tables is 44 in WikiSQL. The scheme of the model is shown in Fig. 2A).

\[D_t = \text{LSTM}(P_{t-1}, (h_{t-1}, c_{t-1}))\]
\[h_0 = (\mathcal{V} \mathcal{W} H([\text{CLS}]), \mathcal{V} d)\]
\[c_0 = (\mathcal{V} \mathcal{W} H([\text{CLS}]), d:2d)\]
\[s_t(i) = \mathcal{W}(\mathcal{V} H + \mathcal{W} D_t)\]
\[p_t(i) = \text{softmax}(s_t(i)).\]

\(P_{t-1}\) stands for the one-hot vector (pointer) at time \(t - 1\), \(h_{t-1}\) and \(c_{t-1}\) are hidden- and cell-vectors of LSTM decoder, \(d\) is the hidden dimension of BERT, \(H_i\) is the BERT output of \(i\)-th token, and \(p_t(i)\) is the probability observing \(i\)-th token at time \(t\).

### 4.3 NL2SQL-LAYER (SQLOVA)

Compared to Shallow-Layer and Decoder-Layer, NL2SQL-Layer contains both encoders and decoders treating the output of table-aware BERT encoder as word-embedding vectors. Like Shallow-Layer, a syntax-guided sketch is adopted and generates SQL query using six separate modules. The resulting model structure resembles SQLNet (Xu et al., 2017) with following differences. Unlike (Xu et al., 2017), our modules do not share parameters. Instead of using pointer network for inferring the where condition values, we train for inferring the start and the end positions of the utterance, following (Dong and Lapata, 2018). Furthermore, the inference of start and end tokens in where-value module depend on both selected where-column and where-operators while the inference relies on where-columns only in (Xu et al., 2017).

Another small difference is that when combining two vectors containing the information about question and headers respectively, a concatenation is used instead of an addition. The scheme of the model is shown in Fig. 2C and the details can be checked from the source code (https://github.com/naver/sqlova).
4.4 Execution-guided decoding.

In decoding (SQL query generation) stage, non-executable (partial) SQL queries are excluded from the output candidates following the strategy suggested in (Wang et al., 2018a). In select clause, (select column, aggregation operator) pairs are excluded when the string-type columns are paired with numerical aggregation operators such as MAX, MIN, SUM, or AVG. The pair with highest joint probability are selected from remaining pairs. In where clause decoding, the executability of each (where column, operator, value) pair is tested by checking the answer returned by the partial SQL query `select agg(col) where colw op val`. Here, col is the predicted select-column, agg is the predicted aggregation operator, colw is one of the where-column candidates, op is where operator, and val stands for the where condition value. If the query returns an empty answer, it is also excluded from the candidates. The final output of where clause is determined by selecting the output maximizing the joint probability estimated from the output of where-number, where-column, where-operator, and where-value modules.

5 Experiment

Pre-trained BERT models (BERT-Large-Uncased\(^1\)) are loaded and fine-tuned with ADAM optimizer with learning rate \(10^{-5}\) whereas the sequence-to-SQL module of NL2SQL-LAYER(SQLOVA) and the decoder in DECODER-LAYER are trained with \(10^{-3}\) learning rate with \(\beta_1 = 0.9, \beta_2 = 0.999\). Batch size is set to 32. To find word vectors, natural language utterance is first tokenized by using Standford CoreNLP (Manning et al., 2014). Each token is further tokenized (into subword level) by WordPiece tokenizer (Devlin et al., 2018; Wu et al., 2016). The headers of the tables and SQL vocabulary in DECODER-LAYER are tokenized by WordPiece tokenizer directly. The PyTorch version of BERT code \(^2\) is used for word embedding and part of the code in NL2SQL-LAYER is influenced by the original SQLNet source code \(^3\). All computations were done on NAVER Smart Machine Learning (NSML) platform (Sung et al., 2017; Kim et al., 2018).

5.1 Accuracy measurement

The logical form (LX) and the execution accuracy (X) on dev set (consisting of 8,421 queries) and test set (consisting of 15,878 queries) of WikiSQL of several models are shown in Table 2. The execution accuracy is measured by evaluating on the answer returned by ‘executing’ the query on the SQL database. The order of where conditions is ignored in measuring logical form accuracy in our models. The top rows in Table 2 show models without execution guidance, and the bottom rows show models augmented with execution-guided decoding (EG). All of our models outperform previous baselines by a large margins. For non-EG scenario, SHALLOW-LAYER shows +5.5% LF and +3.1% X, DECODER-LAYER shows +4.4% LF and +1.8% X, and NL2SQL-LAYER shows +5.3% LF and +2.5% X. For EG case, SHALLOW-LAYER shows +6.4% LF and +0.4% X, DECODER-LAYER shows +7.8% LF and +2.5% X, and NL2SQL-LAYER shows +8.2% LF and +2.5% X.

Interestingly, the performance between our models is SHALLOW-LAYER \(\succeq\) NL2SQL-LAYER > DECODER-LAYER whereas with execution guidance it becomes, NL2SQL-LAYER \(\succeq\) DECODER-LAYER > SHALLOW-LAYER leading us to call NL2SQL-LAYER as SQLOVA together with table-aware BERT encoder due to its overall superiority.

5.2 Accuracy of each module

To understand the performance of SHALLOW-LAYER and NL2SQL-LAYER in detail, the logical form accuracy of each sub-module was calculated and shown in Table. 3. All sub-modules show \(\succeq 95\%\) in accuracy except select-aggregation module. Further investigation of the origin of the low accuracy

\(^1\)https://github.com/google-research/bert
\(^2\)https://github.com/huggingface/pytorch-pretrained-BERT
\(^3\)https://github.com/xiaojunxu/SQLNet
Table 2: Comparison of various models. Logical from accuracy (LF) and execution accuracy (X) on dev and test set of WikiSQL. “EG” stands for “execution-guided”.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev LF (%)</th>
<th>Dev X (%)</th>
<th>Test LF (%)</th>
<th>Test X (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (Zhong et al., 2017)</td>
<td>23.3</td>
<td>37.0</td>
<td>23.4</td>
<td>35.9</td>
</tr>
<tr>
<td>Seq2SQL (Zhong et al., 2017)</td>
<td>49.5</td>
<td>60.8</td>
<td>48.3</td>
<td>59.4</td>
</tr>
<tr>
<td>SQLNet (Xu et al., 2017)</td>
<td>63.2</td>
<td>69.8</td>
<td>61.3</td>
<td>68.0</td>
</tr>
<tr>
<td>PT-MAML (Huang et al., 2018)</td>
<td>63.1</td>
<td>68.3</td>
<td>62.8</td>
<td>68.0</td>
</tr>
<tr>
<td>TypeSQL (Yu et al., 2018)</td>
<td>68.0</td>
<td>74.5</td>
<td>66.7</td>
<td>73.5</td>
</tr>
<tr>
<td>Coarse2Fine (Dong and Lapata, 2018)</td>
<td>72.5</td>
<td>79.0</td>
<td>71.7</td>
<td>78.5</td>
</tr>
<tr>
<td>MQAN (McCann et al., 2018)</td>
<td>76.1</td>
<td>82.0</td>
<td>75.4</td>
<td>81.4</td>
</tr>
<tr>
<td>Annotated Seq2seq (Wang et al., 2018b)</td>
<td>72.1</td>
<td>82.1</td>
<td>72.1</td>
<td>82.2</td>
</tr>
<tr>
<td>IncSQL (Shi et al., 2018)</td>
<td>49.9</td>
<td>84.0</td>
<td>49.9</td>
<td>83.7</td>
</tr>
<tr>
<td><strong>SHALLOW-LAYER (ours)</strong></td>
<td><strong>81.5 (+5.4)</strong></td>
<td><strong>87.4 (+3.2)</strong></td>
<td><strong>80.9 (+5.5)</strong></td>
<td><strong>86.8 (+3.1)</strong></td>
</tr>
<tr>
<td><strong>DECODER-LAYER (ours)</strong></td>
<td><strong>79.7 (+3.6)</strong></td>
<td><strong>85.5 (+1.1)</strong></td>
<td><strong>79.8 (+4.4)</strong></td>
<td><strong>85.5 (+1.8)</strong></td>
</tr>
<tr>
<td><strong>NL2SQL-LAYER (SQLOVA, ours)</strong></td>
<td><strong>81.6 (+5.5)</strong></td>
<td><strong>87.2 (+3.2)</strong></td>
<td><strong>80.7 (+5.5)</strong></td>
<td><strong>86.2 (+2.5)</strong></td>
</tr>
<tr>
<td>PointSQL-EG (Wang et al., 2018a)</td>
<td>67.5</td>
<td>78.4</td>
<td>67.9</td>
<td>78.3</td>
</tr>
<tr>
<td>Coarse2Fine-EG (Wang et al., 2018a)</td>
<td>76.0</td>
<td>84.0</td>
<td>75.4</td>
<td>83.8</td>
</tr>
<tr>
<td>IncSQL-EG (Shi et al., 2018)</td>
<td>51.3</td>
<td>87.2</td>
<td>51.1</td>
<td>87.1</td>
</tr>
<tr>
<td><strong>SHALLOW-LAYER-EG (ours)</strong></td>
<td><strong>82.3 (+6.3)</strong></td>
<td><strong>88.1 (+0.9)</strong></td>
<td><strong>81.8 (+6.4)</strong></td>
<td><strong>87.5 (+0.4)</strong></td>
</tr>
<tr>
<td><strong>DECODER-LAYER-EG (ours)</strong></td>
<td><strong>83.4 (+7.4)</strong></td>
<td><strong>89.9 (+2.7)</strong></td>
<td><strong>83.2 (+7.8)</strong></td>
<td><strong>89.6 (+2.5)</strong></td>
</tr>
<tr>
<td><strong>NL2SQL-LAYER-EG (SQLOVA-EG, ours)</strong></td>
<td><strong>84.2 (+8.2)</strong></td>
<td><strong>90.2 (+3.0)</strong></td>
<td><strong>83.6 (+8.2)</strong></td>
<td><strong>89.6 (+2.5)</strong></td>
</tr>
<tr>
<td>Human performance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>88.2</td>
</tr>
</tbody>
</table>

1 Source code is not opened.
2 Execution guided decoding is employed.
3 Measured over 1,533 randomly chosen samples from WikiSQL test set (Section-6).

reveals that our model effectively achieves the upper bound of WikiSQL task as described in Section 6.

5.3 Ablation study

To understand the importance of each part of the NL2SQL-LAYER, we performed ablation study (Table 4). The results show that in NL2SQL-LAYER, table-aware encoding (without fine-tuning) contributes to overall accuracy by 4.1% (dev) and 3.9% (test) (3rd and 4th rows) which is similar to the results observed in ref. (Dong and Lapata, 2018) where the 3.1% increases observed. This may indicate that approximately 1% contribution is from the use of pre-trained BERT without fine-tuning instead of the GloVe as an word-embedding module. But unlike GloVe, where fine tuning increases only a few percent in accuracy of sub-module (Xu et al., 2017), fine-tuning of BERT increases the accuracy by 11.7% (dev) and 12.2% (test) which may be attributed to the use of many-layers of self-attentions (Vaswani et al., 2017). Use of BERT-Base decreases the accuracy by 1.3% on both dev and test set compared to BERT-Large cases.

6 Analysis

6.1 Error Analysis on the WikiSQL Dataset

There are 1,533 mismatches in logical form between the ground-truth and the predictions from SQLOVA in WikiSQL dev set consisting of 8,421 examples. To investigate the origin of the mismatches, 100 samples were randomly selected from 1,533 examples and analyzed manually. Interestingly, 30 out of 100 examples are not answerable under WikiSQL task. We categorize 30 unanswerable examples into two main types:

- **Type I.** Questions are not answerable with given information (questions and table headers). For example, fourth sample of the Table. 6 in Appendix (QID-3050) shows table contents are required. Other representative examples are QID-1986, and QID-261.

- **Type II.** Questions are generated from incorrectly retrieved tables. For example,
Table 3: The logical form accuracy of each sub-module over WikiSQL dev set. s-col, s-agg, w-num, w-col, w-op and w-val stand for select-column, select-aggregation, where-number, where-column, where-operator, and where-value respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>s-col</th>
<th>s-agg</th>
<th>w-num</th>
<th>w-col</th>
<th>w-op</th>
<th>w-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQLOVA, Dev</td>
<td>97.3</td>
<td>90.5</td>
<td>98.7</td>
<td>94.7</td>
<td>97.5</td>
<td>95.9</td>
</tr>
<tr>
<td>SQLOVA, Test</td>
<td>96.8</td>
<td>90.6</td>
<td>98.5</td>
<td>94.3</td>
<td>97.3</td>
<td>95.4</td>
</tr>
<tr>
<td>SHALLOW-LAYER, Dev</td>
<td>97.2</td>
<td>89.8</td>
<td>98.8</td>
<td>95.3</td>
<td>97.8</td>
<td>96.0</td>
</tr>
<tr>
<td>SHALLOW-LAYER, Test</td>
<td>97.0</td>
<td>89.8</td>
<td>98.6</td>
<td>94.8</td>
<td>97.6</td>
<td>95.8</td>
</tr>
</tbody>
</table>

Table 4: The results of ablation study. Logical form accuracy (LF) and execution accuracy (X) on dev and test sets of WikiSQL are shown.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev LF (%)</th>
<th>Dev X (%)</th>
<th>Test LF (%)</th>
<th>Test X (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQLOVA</td>
<td>81.6</td>
<td>87.2</td>
<td>80.7</td>
<td>86.2</td>
</tr>
<tr>
<td>− BERT-Large + BERT-Base</td>
<td>80.3</td>
<td>85.8</td>
<td>79.4</td>
<td>85.2</td>
</tr>
<tr>
<td>− Fine-tuning</td>
<td>69.9</td>
<td>77.0</td>
<td>68.5</td>
<td>75.6</td>
</tr>
<tr>
<td>− BERT-Large + GloVe</td>
<td>65.8</td>
<td>72.9</td>
<td>64.6</td>
<td>71.7</td>
</tr>
</tbody>
</table>

QID-3328 shows incorrect header names making answering question impossible.

Further analysis over remaining 70 answerable examples shows that there are 51 incorrect logical forms in ground truth (e.g. QID-5611, 2159). Interestingly, among 51 examples, 44 logical forms are correctly predicted by SQLOVA indicating that the actual performance of proposed models could be higher than the accuracies shown in Table 2. This also may imply that most of training examples in WikiSQL have correct ground truth for the proper training of models. The results are summarized in Table 5. 100 samples are presented in Table 6 in Appendix with the type of errors. The results led us to conjecture 90% accuracy could be the near upper bound in WikiSQL task for answerable and non erroneous questions when table-contents are not available.

Some of the errors in WikiSQL dataset could be introduced while paraphrasing queries generated automatically from the templates (Zhong et al., 2017). As questions were generated without considering the table contents, paraphrasing could change the meanings of queries especially when the quantitative answer is required. For example, QID-40 in Table 6 is related to an “year” and the ground truth SQL query includes unnecessary COUNT aggregation operators.

Table 5: Contingency table of 70 answerable questions. Corresponding 70 ground truth (GT) and predicted-SQL queries by from SQLOVA were manually separated to correct and incorrect cases.

<table>
<thead>
<tr>
<th>SQL (GT)</th>
<th>correct</th>
<th>incorrect</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>44</td>
<td>0</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>total</td>
<td>51</td>
<td>19</td>
<td>70</td>
</tr>
</tbody>
</table>

6.2 Human performance

The human performance on WikiSQL dataset has not been measured despite of its popularity. Here, we provide the approximate human performance by collecting answers from 246 different crowdworkers through Amazon Mechanical Turk over 1,533 randomly sampled data from the WikiSQL test set composed of 15,878 samples. During the evaluation, crowdworkers were asked either to find value(s) or to compute a value using the given table for given questions, as in the same condition for models to measure execution accuracy in the WikiSQL task. The execution accuracy of crowdworkers on this randomly sampled data is 88.2% (Table 2).

Even though only the parts of the WikiSQL test set were used for the performance evaluation, 1,533 samples are enough to approximate the human performance which can be shown by assuming that the probability of having cor-
rect answer for each question from human \( (q) \) is independent and identically distributed. For example, if \( q \sim 0.9 \), the mean number of correct answers \( (\mu) \) is 1.380 with standard deviation \( \sigma \sim 12 \) as it follows binomial distribution. Thus, the expected fluctuation in execution accuracy is \( \sim 0.8\% \) in this case. For \( 0.75 \leq q \leq 0.95 \), the fluctuation in accuracy is less than 1.1%.

Errors made by crowds were similar to models. Mismatch of target columns, condition columns, etc. One notable mistake by humans was ambiguity of natural language that was not considered in models. For example, when NL is asking a target column value with more than two conditions, certain portion of crowdworkers showed tendency to choose a target column value with one condition because multiple conditions were written with “and”, and it is often considered as the meaning of “or” in real life.

6.3 Precision-Recall-Based Performance Measure

The accuracy metric in Table. 2 of WikiSQL task treats predicting no answers as giving incorrect answers. However, the ability to not generate answer when the model is not confident about the prediction is an another important evaluation metric in practice. Here, we consider the output probability of generating SQL query of the models as the confidence score and consider the question is unanswerable when the score is low. With this setting, we performed a precision-recall-based analysis with following four categorization of the results:

- True positive (TP): correct answer with high confidence
- False negative (FN): correct answer with low confidence
- False positive (FP): incorrect answer with high confidence
- True negative (TN): incorrect answer with low confidence

The precision-recall curve and its area under curve value of the proposed model are shown in Fig. 3. The result shows that SQLova assign low probability to wrong predictions. The calculated precision is higher than 95% with 80% recall.

![Figure 3: Precision-Recall curve and area under curve (AUC) with SQLova (blue) and SQLova-EG (orange). Precision and recall rates are controlled by varying the threshold value for the confidence score.](image)

7 Conclusion

In this paper, we demonstrate the effectiveness of table-aware word contextualization on a popular semantic parsing task, WikiSQL. We propose BERT-based table-aware encoder and three task-specific modules with different model complexity on top of the encoder, namely SHALLOW-LAYER, DECODER-LAYER, and NL2SQL-LAYER. We show that even the simplest module, SHALLOW-LAYER, can outperform the previous best model, but a sufficiently dense module, NL2SQL-LAYER, gives the best result across several different settings. We hope our detailed exploration of the models provides an insight on how word contextualization can be considered in other semantic parsing tasks as well. We further show that our models effectively achieve the upper bound of accuracy in WikiSQL task by performing detailed error analysis together with human-performance evaluation.

Acknowledgments

We thank Clova AI members for their great support, especially Jung-Woo Ha for proofreading the manuscript, Sungdong Kim for providing help on using BERT, Guwan Kim for insightful comments, and Dongjun Lee for
the advice in the preparation of the demo. We also thank the Hugging Face Team for sharing the PyTorch implementation of BERT.

References


Po-Sen Huang, Chenglong Wang, Rishabh Singh, Wen tau Yih, and Xiaodong He. 2018. Natural language to structured query generation via meta-learning. In NAACL.


Hanjoo Kim, Minkyu Kim, Dongjoo Seo, Jinwoong Kim, Heungseok Park, Soeun Park, Hyunwoo Jo, KyungHyun Kim, Youngil Yang, Youngkwon Kim, Nako Sung, and Jung-Woo Ha. 2018. NSML: meet the mlaas platform with a real-world case study. CoRR, abs/1810.09957.


Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah,


## Appendix

### A.1 100 Examples in the WikiSQL Dataset

Table 6: The dataset examples from WikiSQL dev set used in Section. 6. 100 samples were randomly selected from 1,533 mismatches between the ground-truth and the predictions of SQL2OVA. QID denotes an index of the question among 8,421 wikiSQL dev set data. There are three types of queries: natural language queries (NL), ground truth SQL queries (SQL (T)), predicted SQL queries (SQL (P)). Other fields indicate ground truth answer (ANS (T)), predicted answer (ANS (P)), and a type of error (ERROR).

<table>
<thead>
<tr>
<th>No.</th>
<th>QID</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3650</td>
<td>NL</td>
<td>How many Byes have Against of 1076 and Wins smaller than 13?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (T)</td>
<td>SELECT avg(Byes) FROM 2-1552908-21 WHERE Against = 1076 AND Wins &lt; 13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (P)</td>
<td>SELECT count(Byes) FROM 2-1552908-21 WHERE Wins &lt; 13 AND Against = 1076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (T)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (P)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERROR</td>
<td>Ground Truth</td>
</tr>
<tr>
<td>2</td>
<td>2090</td>
<td>NL</td>
<td>What year was mcmahon stadium founded?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (T)</td>
<td>SELECT max(Founded) FROM 1-27599216-6 WHERE Football stadium = McMa- hon Stadium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (P)</td>
<td>SELECT (Founded) FROM 1-27599216-6 WHERE Football stadium = mcmahon stadium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (T)</td>
<td>1966.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (P)</td>
<td>1966.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERROR</td>
<td>Ground Truth</td>
</tr>
<tr>
<td>3</td>
<td>7062</td>
<td>NL</td>
<td>What is the Mintage after 2006 of the Ruby-Throated Hummingbird Theme coin?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (T)</td>
<td>SELECT max(Mintage) FROM 2-17757354-2 WHERE Year &gt; 2006 AND Theme = ruby-throated hummingbird</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (P)</td>
<td>SELECT (Mintage) FROM 2-17757354-2 WHERE Year &gt; 2006 AND Theme = ruby-throated hummingbird</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (T)</td>
<td>25,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (P)</td>
<td>25,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERROR</td>
<td>Ground Truth</td>
</tr>
<tr>
<td>4</td>
<td>3050</td>
<td>NL</td>
<td>Which team is in the Southeast with a home at Philips Arena?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TBL</td>
<td>“Conference”, “Division”, “Team”, “City”, “Home Arena”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (T)</td>
<td>SELECT (Team) FROM 2-14519555-8 WHERE Division = southeast AND Home Arena = philips arena</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL (P)</td>
<td>SELECT (Team) FROM 2-14519555-8 WHERE Conference = southeast AND Home Arena = philips arena</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (T)</td>
<td>atlanta hawks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANS (P)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERROR</td>
<td>Question</td>
</tr>
<tr>
<td>5</td>
<td>795</td>
<td>NL</td>
<td>If the equation is (10 times 8) + 4, what would be the 2nd throw?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TBL</td>
<td>“1st throw”, “2nd throw”, “3rd throw”, “Equation”, “Result”</td>
</tr>
</tbody>
</table>
### SQL Queries and Answers

#### 6. How many times has Ma Long won the men’s singles?

| SQL (T) | `SELECT max(2nd throw) FROM 1-17265535-6 WHERE Equation = (10 times 8) + 4` |
| SQL (P) | `SELECT (2nd throw) FROM 1-17265535-6 WHERE Equation = (10 times 8) + 4` |
| Ans (T) | 4.0 |
| Ans (P) | 4.0 |

#### 7. What was the score between Marseille and Manchester United on the second leg of the Champions League Round of 16?

| SQL (T) | `SELECT (2nd leg score**) FROM 1-26910311-8 WHERE Opponent = Marseille` |
| SQL (P) | `SELECT (2nd leg score**) FROM 1-26910311-8 WHERE Team = marseille AND Contest and round = champions league round of 16 AND Opponent = manchester united` |
| Ans (T) | 2–1 (h) |
| Ans (P) | None |

#### 8. How many incumbents come from Alvin Bush’s district?

| SQL (T) | `SELECT count(Candidates) FROM 1-1341930-38 WHERE Incumbent = Alvin Bush` |
| SQL (P) | `SELECT count(Incumbent) FROM 1-1341930-38 WHERE District = alvin bush` |
| Ans (T) | 1 |
| Ans (P) | 0 |

#### 9. What’s the total of rank 8 when Silver medals are 0 and gold is more than 1?

| SQL (T) | `SELECT count(Total) FROM 2-18807607-2 WHERE Silver = 0 AND Rank = 8 AND Gold > 1` |
| SQL (P) | `SELECT sum(Total) FROM 2-18807607-2 WHERE Rank = 8 when silver medals are 0 AND Gold > 1 AND Silver = 0` |
| Ans (T) | 0 |
| Ans (P) | None |

#### 10. How many millions of U.S. viewers watched the episode “Buzzkill”?

| SQL (T) | `SELECT count(U.S. viewers (millions)) FROM 1-12570759-2 WHERE Title = "Buzzkill"` |
| SQL (P) | `SELECT count(U.S. viewers (millions)) FROM 1-12570759-2 WHERE Title = "Buzzkill"` |
| Ans (T) | None |
| Ans (P) | None |
11 3770 NL What is the total number of offensive rebounds for players with under 65 total rebounds, 5 defensive rebounds, and under 7 assists?


SQL (T) SELECT count(Off Reb) FROM 2-15746812-4 WHERE Total Reb < 65 AND Def Reb = 5 AND Asst < 7

SQL (P) SELECT count(Asst) FROM 2-15746812-4 WHERE Off Reb = 5 defensive rebounds AND Total Reb < 65 AND Asst < 7

ANS (T) 0
ANS (P) 0

ERROR Both Ground Truth and Question

12 6927 NL Which Number of electorates (2009) has a Constituency number of 46?

TBL "Constituency number", "Name", "Reserved for ( SC / ST /None)", "District", "Number of electorates (2009)"

SQL (T) SELECT avg(Number of electorates (2009)) FROM 2-17922541-1 WHERE Constituency number = 46

SQL (P) SELECT (Number of electorates (2009)) FROM 2-17922541-1 WHERE Constituency number = 46

ANS (T) 136.0
ANS (P) 136,987

ERROR Ground Truth

13 261 NL Name the perfect stem for jo

TBL "Perfect stem", "Future stem", "Imperfect stem", "Short stem", "Meaning"

SQL (T) SELECT count(Perfect stem) FROM 1-12784134-24 WHERE Short stem = jo

SQL (P) SELECT (Perfect stem) FROM 1-12784134-24 WHERE Imperfect stem = jo

ANS (T) 1
ANS (P) None

ERROR Both Ground Truth and Question

14 5611 NL Who was home at Princes Park?

TBL "Home team", "Home team score", "Away team", "Away team score", "Venue", "Crowd", "Date"

SQL (T) SELECT (Home team score) FROM 2-10809157-18 WHERE Venue = princes park

SQL (P) SELECT (Home team) FROM 2-10809157-18 WHERE Venue = princes park

ANS (T) 9.16 (70)
ANS (P) fitzroy

ERROR Ground Truth

15 1978 NL How many games were played where the height of the player is 1.92m?

TBL "Player", "Position", "Starting No.#", "D.O.B", "Club", "Height", "Weight", "Games"

SQL (T) SELECT count(Games) FROM 1-26847237-2 WHERE Height = 1.92m

SQL (P) SELECT (Games) FROM 1-26847237-2 WHERE Height = 1.92m

ANS (T) 1
ANS (P) 7.0
16 2159 NL Which game had a score of W 95-85?

TBL  “Game”, “Date”, “Team”, “Score”, “High points”, “High rebounds”, “High assists”, “Location Attendance”, “Record”

SQL (T) SELECT min(Game) FROM 1-27902171-7 WHERE Score = W 95-85
SQL (P) SELECT (Game) FROM 1-27902171-7 WHERE Score = w 95-85
ANS (T) 48.0
ANS (P) 48.0

ERROR Ground Truth

17 2812 NL What is the name of the driver with a rotax max engine, in the rotax heavy class, with arrow as chassis and on the TWR Raceline Seating team?

TBL  “Team”, “Class”, “Chassis”, “Engine”, “Driver”

SQL (T) SELECT (Driver) FROM 2-15162596-2 WHERE Engine = rotax max AND Class = rotax heavy AND Chassis = arrow AND Team = twr raceline seating
SQL (P) SELECT (Driver) FROM 2-15162596-2 WHERE Team = twr raceline seating AND Class = rotax heavy AND Chassis = arrow AND Engine = rotax max engine, in the rotax heavy

ANS (T) rod clarke
ANS (P) None

ERROR Ground Truth

18 841 NL When there was a bye in the round of 32, what was the result in the round of 16?


SQL (T) SELECT (Semifinals) FROM 1-1745820-5 WHERE Round of 32 = Bye
SQL (P) SELECT (Round of 16) FROM 1-1745820-5 WHERE Round of 32 = bye

ANS (T) did not advance
ANS (P) simelane (swz) w(rsc)

ERROR Ground Truth

19 5456 NL What year has a Schwante smaller than 2.043, an Eichstädt smaller than 848, and a Bärenklau smaller than 1.262?


SQL (T) SELECT count(Year) FROM 2-11680175-1 WHERE Schwante < 2.043 AND Eichstädt < 848 AND Bärenklau < 1.262
SQL (P) SELECT sum(Year) FROM 2-11680175-1 WHERE Schwante < 2.043 AND Bärenklau < 1.262 AND Eichstädt < 848

ANS (T) 0
ANS (P) None

ERROR Ground Truth

20 3328 NL What was the total in 2009 for years of river vessels when 2008 was more than 8,030 and 2007 was more than 1,411,414?


SQL (T) SELECT count(2009) FROM 2-13823555-1 WHERE Years = river vessels AND 2007 > 1,411,414 AND 2008 > 8,030
SQL (P) SELECT sum(2009) FROM 2-13823555-1 WHERE Years = river vessels AND 2007 > 1,411,414 AND 2008 > 8,030

ANS (T) 1
ANS (P) 6.0

ERROR Both Ground Truth and Question
21 8111 NL What is the language of the film Rosie?
TBL “Year (Ceremony)”, “Film title used in nomination”, “Original title”, “Language(s)”, “Result”
SQL (T) SELECT (Language(s)) FROM 2-13330057-1 WHERE Original title = rosie
SQL (P) SELECT (Language(s)) FROM 2-13330057-1 WHERE Film title used in nomination = rosie
ANS (T) dutch
ANS (P) dutch
ERROR Question

22 2323 NL What team hired Renato Gaúcho?
TBL “Team”, “Outgoing manager”, “Manner of departure”, “Date of vacancy”, “Position in table”, “Replaced by”, “Date of appointment”
SQL (T) SELECT (Team) FROM 1-29414946-3 WHERE Replaced by = Renato Gaúcho
SQL (P) SELECT (Team) FROM 1-29414946-3 WHERE Outgoing manager = renato gaúcho
ANS (T) atlético paranaense
ANS (P) grêmio
ERROR None

23 156 NL What is the number of the player who went to Southern University?
TBL “Player”, “No.(s)”, “Height in Ft.”, “Position”, “Years for Rockets”, “School/Club Team/Country”
SQL (T) SELECT (No.(s)) FROM 1-11734041-9 WHERE School/Club Team/Country = Southern University
SQL (P) SELECT count(No.(s)) FROM 1-11734041-9 WHERE School/Club Team/Country = southern university
ANS (T) 6
ANS (P) 1
ERROR Question

24 7725 NL How many cuts made in the tournament he played 13 times?
SQL (T) SELECT sum(Cuts made) FROM 2-12702607-1 WHERE Events > 13
SQL (P) SELECT (Cuts made) FROM 2-12702607-1 WHERE Wins = 13
ANS (T) None
ANS (P) None
ERROR Question

25 19 NL How many capital cities does Australia have?
TBL “Country ( exonym )”, “Capital ( exonym )”, “Country ( endonym )”, “Capital ( endonym )”, “Official or native language(s) (alphabet/script)”
SQL (T) SELECT count(Capital ( endonym )) FROM 1-1008653-1 WHERE Country ( endonym ) = Australia
SQL (P) SELECT count(Capital ( exonym )) FROM 1-1008653-1 WHERE Country ( exonym ) = australia
ANS (T) 1
ANS (P) 1
ERROR Question

26 1591 NL What was the rating for Brisbane the week that Adelaide had 94000?
27 7106 NL What was the score of the BCS National Championship game?

TBL “Date”, “Bowl Game”, “Big Ten Team”, “Opp. Team”, “Score”

SQL (T) SELECT (Score) FROM 2-18102742-1 WHERE Bowl Game = bcs national championship

SQL (P) SELECT (Score) FROM 2-18102742-1 WHERE Bowl Game = bcs national championship game

ANS (T) 38-24

ANS (P) None

ERROR None

28 6440 NL In the match where the away team scored 2.7 (19), how many people were in the crowd?

TBL “Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”

SQL (T) SELECT max(Crowd) FROM 2-10790397-5 WHERE Away team score = 2.7 (19)

SQL (P) SELECT (Crowd) FROM 2-10790397-5 WHERE Away team score = 2.7 (19)

ANS (T) 15,000

ANS (P) 15,000

ERROR Ground Truth

29 5707 NL What is Fitzroy’s Home team Crowd?

TBL “Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”

SQL (T) SELECT sum(Crowd) FROM 2-10809142-16 WHERE Home team = fitzroy

SQL (P) SELECT (Crowd) FROM 2-10809142-16 WHERE Home team = fitzroy

ANS (T) 20.0

ANS (P) 20,000

ERROR Ground Truth

30 5055 NL What is the To par of Player Andy North with a Total larger than 153?

TBL “Player”, “Country”, “Year(s) won”, “Total”, “To par”

SQL (T) SELECT count(To par) FROM 2-17162255-3 WHERE Player = andy north AND Total > 153

SQL (P) SELECT (To par) FROM 2-17162255-3 WHERE Player = andy north AND Total > 153

ANS (T) 0

ANS (P) None

ERROR None

31 4785 NL What is the name of the free transfer fee with a transfer status and an ENG country?

TBL “Name”, “Country”, “Status”, “Transfer window”, “Transfer fee”

SQL (T) SELECT (Name) FROM 2-16549823-7 WHERE Transfer fee = free AND Status = transfer AND Country = eng

SQL (P) SELECT (Name) FROM 2-16549823-7 WHERE Country = eng AND Status = AND Transfer fee = free transfer

ANS (T) bailey
32 1199 NL what is the No in series when Rob Wright & Debra J. Fisher & Erica Messer were the writers?

TBL “No. in series”, “No. in season”, “Title”, “Directed by”, “Written by”, “Original air date”, “Production code”, “U.S. viewers (millions)”

SQL (T) SELECT min(No. in series) FROM 1-21313327-1 WHERE Written by = Rob Wright & Debra J. Fisher & Erica Messer

SQL (P) SELECT (No. in series) FROM 1-21313327-1 WHERE Written by = rob wright & debra j. fisher & erica messer

ANS (T) 149.0

ANS (P) 149.0

ERROR None

33 7912 NL When did Gaspare Bona win the Pozzo Circuit?


SQL (T) SELECT (Date) FROM 2-12631771-2 WHERE Winning driver = gaspare bona AND Name = pozzo circuit

SQL (P) SELECT (Date) FROM 2-12631771-2 WHERE Circuit = pozzo AND Winning driver = gaspare bona

ANS (T) 20 march

ANS (P) 20 march

ERROR Ground Truth

34 3602 NL In what Year did the German Open have Yoo Sang-Hee as Partner?

TBL “Outcome”, “Event”, “Year”, “Venue”, “Partner”

SQL (T) SELECT (Year) FROM 2-14895591-2 WHERE Partner = yoo sang-hee AND Venue = german open

SQL (P) SELECT (Year) FROM 2-14895591-2 WHERE Event = german open AND Partner = yoo sang-hee

ANS (T) 1986

ANS (P) None

ERROR Question

35 2223 NL What daft pick number is the player coming from Regina Pats (WHL)?

TBL “Pick #”, “Player”, “Position”, “Nationality”, “NHL team”, “College/junior/club team”

SQL (T) SELECT (Pick #) FROM 1-2850912-1 WHERE College/junior/club team = Regina Pats (WHL)

SQL (P) SELECT min(Pick #) FROM 1-2850912-1 WHERE College/junior/club team = regina pats (whl)

ANS (T) 21.0

ANS (P) 21.0

ERROR None

36 738 NL Name the location for illinois


SQL (T) SELECT (Location) FROM 1-1672976-7 WHERE Big Ten Team = Illinois

SQL (P) SELECT (Location) FROM 1-1672976-7 WHERE ACC Team = illinois

ANS (T) littlejohn coliseum • clemson, sc

ANS (P) None
37. WHAT IS THE WEEK WITH AN ATTENDANCE OF 75,555?


SQL (T): SELECT sum(Week) FROM 2-16764708-1 WHERE Attendance = 75,555

SQL (P): SELECT (Week) FROM 2-16764708-1 WHERE Attendance = 75,555

ANS (T): 11.0

ANS (P): 11.0

38. What was the attendance of the game that had an away team of FK Mogren?


SQL (T): SELECT (Attendance) FROM 2-13883437-1 WHERE Guest = FK Mogren

SQL (P): SELECT (Attendance) FROM 2-13883437-1 WHERE Home = away

ANS (T): 1.2

ANS (P): None

39. What’s Brazil’s lane with a time less than 21.15?

TBL: “Rank”, “Lane”, “Athlete”, “Nationality”, “Time”, “React”

SQL (T): SELECT min(Lane) FROM 2-18569011-6 WHERE Nationality = Brazil AND Time < 21.15

SQL (P): SELECT sum(Lane) FROM 2-18569011-6 WHERE Nationality = Brazil AND Time < 21.15

ANS (T): None

ANS (P): None

40. When did Hans Hartmann drive?


SQL (T): SELECT count(Year) FROM 2-14287417-3 WHERE Driver = Hans Hartmann

SQL (P): SELECT (Year) FROM 2-14287417-3 WHERE Driver = Hans Hartmann

ANS (T): 1

ANS (P): 1939.0

41. How many ties did he have when he had 1 penalties and more than 20 conversions?


SQL (T): SELECT sum(Drawn) FROM 2-1828549-1 WHERE Penalties = 1 AND Conversions > 20

SQL (P): SELECT (Drawn) FROM 2-1828549-1 WHERE Conversions > 20 AND Penalties = 1

ANS (T): None

ANS (P): None

42. What position does the player from Arkansas play?


SQL (T): SELECT (Pos.) FROM 2-11482079-13 WHERE School/Country = Arkansas

SQL (P): SELECT (Pos.) FROM 2-11482079-13 WHERE School/Country = Arkansas

ANS (T): c
43 3314 NL What is the lowest number of bronze a short track athlete with 0 gold medals has?


SQL (T) SELECT min(Bronze) FROM 2-13554889-6 WHERE Sport = short track AND Gold = 0

SQL (P) SELECT min(Bronze) FROM 2-13554889-6 WHERE Type = short track AND Gold < 0

ANS (T) 2.0
ANS (P) None
ERROR None

44 212 NL What is the toll for heavy vehicles with 3/4 axles at Verkeerdevlei toll plaza?

TBL "Name", "Location", "Light vehicle", "Heavy vehicle (2 axles)", "Heavy vehicle (3/4 axles)", "Heavy vehicle (5+ axles)"

SQL (T) SELECT (Heavy vehicle (3/4 axles)) FROM 1-1211545-2 WHERE Name = Verkeerdevlei Toll Plaza

SQL (P) SELECT (Heavy vehicle (3/4 axles)) FROM 1-1211545-2 WHERE Heavy vehicle (3/4 axles) = verkeerdevlei toll plaza

ANS (T) r117.00
ANS (P) None
ERROR None

45 6779 NL What week was the opponent the San Diego Chargers?

TBL "Week", "Date", "Opponent", "Result", "Kickoff Time", "Attendance"

SQL (T) SELECT avg(Week) FROM 2-17643221-2 WHERE Opponent = san diego chargers

SQL (P) SELECT (Week) FROM 2-17643221-2 WHERE Opponent = san diego chargers

ANS (T) 1.0
ANS (P) 1.0
ERROR Ground Truth

46 1089 NL Name the total number of date for 163-77

TBL "Game", "Date", "Opponent", "Score", "High points", "High rebounds", "High assists", "Location", "Record"

SQL (T) SELECT count(Date) FROM 1-19789597-5 WHERE Score = L 63-77

SQL (P) SELECT count(Date) FROM 1-19789597-5 WHERE Record = L 63-77

ANS (T) 1
ANS (P) 0
ERROR Question

47 3499 NL Which driver has less than 37 wins and at 14.12%?

TBL "Driver", "Seasons", "Entries", "Wins", "Percentage"

SQL (T) SELECT avg(Entries) FROM 2-13599687-6 WHERE Wins < 37 AND Percentage = 14.12%

SQL (P) SELECT (Driver) FROM 2-13599687-6 WHERE Wins < 37 AND Percentage = 14.12%

ANS (T) 177.0
ANS (P) niki lauda
ERROR Ground Truth

48 1035 NL Name the number of candidates for # of seats won being 43
**50 2286 NL** What is the area when the Iga name is Ahoada East?

TBL “LGA Name”, “Area (km 2 )”, “Census 2006 population”, “Administrative capital”, “Postal Code”

SQL (T) SELECT max(Area (km 2 )) FROM 1-28891101-3 WHERE LGA Name = Ahoada East

SQL (P) SELECT (Area (km 2 )) FROM 1-28891101-3 WHERE LGA Name = ahoada east

ANS (T) 341.0

ANS (P) 341.0

ERROR Ground Truth

**51 4561 NL** How much Overall has a Name of bob anderson?

TBL “Round”, “Pick”, “Overall”, “Name”, “Position”, “College”

SQL (T) SELECT count(Overall) FROM 2-17100961-17 WHERE Name = bob anderson

SQL (P) SELECT sum(Overall) FROM 2-17100961-17 WHERE Name = bob anderson

ANS (T) 1

ANS (P) 68.0

ERROR Ground Truth

**52 3925 NL** What is the value for the item “Tries” when the value of the item “Played” is 18 and the value of the item “Points” is 375?


SQL (T) SELECT (Tries For) FROM 2-15467476-4 WHERE Played = 18 AND Points against = 375

SQL (P) SELECT (Tries Against) FROM 2-15467476-4 WHERE Played = 18 AND Points for = 375

ANS (T) 54

ANS (P) None

ERROR Question

**53 1851 NL** Which year did enrolled Gambier members leave?


SQL (T) SELECT min(Left) FROM 1-261946-3 WHERE Location (all in Ohio) = Gambier

SQL (P) SELECT (Left) FROM 1-261946-3 WHERE Nickname = gambier
54 7182 NL What is the finishing time with a 2/1q finish on the Meadowlands track?
SQL (T) SELECT (Fin. Time) FROM 2-18744745-2 WHERE Finish = 2/1q AND Track = the meadowlands
SQL (P) SELECT (Fin. Time) FROM 2-18744745-2 WHERE Track = meadowlands AND Finish = 2/1q
ANS (T) 1:47.1
ANS (P) None
ERROR None

55 654 NL What is the total population in the city/town of Arendal?
TBL “City/town”, “Municipality”, “County”, “City/town status”, “Population”
SQL (T) SELECT count(Population) FROM 1-157826-1 WHERE City/town = Arendal
SQL (P) SELECT sum(Population) FROM 1-157826-1 WHERE City/town = arendal
ANS (T) 1
ANS (P) 39826.0
ERROR None

56 5304 NL What is the total avg/g of McCrary, Greg?
SQL (T) SELECT count(Avg/G) FROM 2-16981858-6 WHERE Name = mccrary, greg
SQL (P) SELECT sum(Avg/G) FROM 2-16981858-6 WHERE Name = mccrary, greg
ANS (T) 1
ANS (P) 58.9
ERROR None

57 7438 NL With a Col (m) larger than 2012, what is Mount Kazbek’s Prominence (m)?
TBL “Peak”, “Country”, “Elevation (m)”, “Prominence (m)”, “Col (m)”
SQL (T) SELECT (Prominence (m)) FROM 2-18918776-12 WHERE Col (m) > 2012 AND Peak = mount kazbek
SQL (P) SELECT max(Prominence (m)) FROM 2-18918776-12 WHERE Peak = mount kazbek AND Col (m) > 2012
ANS (T) 2,353
ANS (P) 2,353
ERROR None

58 475 NL Name the finished for kerry katona
SQL (T) SELECT count(Finished) FROM 1-14345690-4 WHERE Celebrity = Kerry Katona
SQL (P) SELECT (Finished) FROM 1-14345690-4 WHERE Celebrity = kerry katona
ANS (T) 1
ANS (P) 1st
ERROR None

59 1449 NL How many games had been played when the Mavericks had a 46-22 record?
TBL “Game”, “Date”, “Team”, “Score”, “High points”, “High rebounds”, “High assists”, “Location Attendance”, “Record”
SQL (T) SELECT max(Game) FROM 1-23284271-9 WHERE Record = 46-22
SQL (P) SELECT (Game) FROM 1-23284271-9 WHERE Record = 46-22
ANS (T) None
ANS (P) None
ERROR Ground Truth
SQL (P)  SELECT (Game) FROM 1-23284271-9 WHERE Record = 46-22
ANS (T)  68.0
ANS (P)  68.0
ERROR  Ground Truth

60  598  NL  Name the best supporting actress for sun honglei for mongol
TBL  “Year”, “Best Film”, “Best Director”, “Best Actor”, “Best Actress”, “Best Supporting Actor”, “Best Supporting Actress”
SQL (T)  SELECT (Best Supporting Actress) FROM 1-15301258-1 WHERE Best Supporting Actor = Sun Honglei for Mongol
SQL (P)  SELECT (Best Supporting Actress) FROM 1-15301258-1 WHERE Best Film = mongol AND Best Actor = sun honglei
ANS (T)  joan chen for the sun also rises
ANS (P)  None
ERROR  Question

61  3730  NL  What is the average year of the Fantasia Section Award?
TBL  “Festival”, “Year”, “Result”, “Award”, “Category”
SQL (T)  SELECT avg(Year) FROM 2-1201864-1 WHERE Award = fantasia section award
SQL (P)  SELECT avg(Year) FROM 2-1201864-1 WHERE Award = fantasia section
ANS (T)  1999.0
ANS (P)  None
ERROR  None

62  5882  NL  What is the average year of the Fantasia Section Award?
TBL  “Festival”, “Year”, “Result”, “Award”, “Category”
SQL (T)  SELECT avg(Year) FROM 2-1201864-1 WHERE Award = fantasia section award
SQL (P)  SELECT avg(Year) FROM 2-1201864-1 WHERE Award = fantasia section
ANS (T)  1999.0
ANS (P)  None
ERROR  None

63  1430  NL  Name the surface for philadelphia
TBL  “Outcome”, “Year”, “Championship”, “Surface”, “Opponent in the final”, “Score in the final”
SQL (T)  SELECT (Surface) FROM 1-23235767-4 WHERE Championship = Philadelphia
SQL (P)  SELECT (Surface) FROM 1-23235767-4 WHERE Opponent in the final = philadelphia
ANS (T)  carpet
ANS (P)  None
ERROR  Question

64  2565  NL  What was attendance of the whole season when the average attendance for League Cup was 32,415?
TBL  “Season”, “Season Total Att.”, “K-League Season Total Att.”, “Regular Season Average Att.”, “League Cup Average Att.”, “FA Cup Total / Average Att.”, “ACL Total / Average Att.”, “Friendly Match Att.”
SQL (T)  SELECT (Season Total Att.) FROM 2-1056336-11 WHERE League Cup Average Att. = 32,415
ANS (T)  35.416666666666664
ANS (P)  None
ERROR  None
**65 2891**

If played is 22 and the tries against are 43, what are the points?

| SQL (T) | SELECT (Points for) FROM 2-13741576-4 WHERE Played = 22 AND Tries against = 43 |
| SQL (P) | SELECT (Points) FROM 2-13741576-4 WHERE Played = 22 AND Tries against = 43 |
| ANS (T) | 353 |
| ANS (P) | 46 |

**ERROR Question**

**66 6533**

Name the Score united states of tom watson in united state?

| TBL | “Place”, “Player”, “Country”, “Score”, “To par” |
| SQL (T) | SELECT (Score) FROM 2-18113463-4 WHERE Country = united states AND Player = tom watson |
| SQL (P) | SELECT (Score) FROM 2-18113463-4 WHERE Place = united states AND Player = tom watson AND Country = united states |
| ANS (T) | 68.0 |
| ANS (P) | None |

**ERROR Question**

**67 7425**

What is galway county’s total?

| TBL | “Rank”, “Player”, “County”, “Tally”, “Total”, “Opposition” |
| SQL (T) | SELECT sum(Total) FROM 2-18936986-3 WHERE County = galway |
| SQL (P) | SELECT (Total) FROM 2-18936986-3 WHERE County = galway |
| ANS (T) | 9.0 |
| ANS (P) | 9.0 |

**ERROR Ground Truth**

**68 7070**

What is the date of the zolder circuit, which had a.z.k./roc-compétition a.z.k./roc-compétition as the winning team?

| TBL | “Round”, “Circuit”, “Date”, “Winning driver”, “Winning team” |
| SQL (T) | SELECT (Date) FROM 2-17997366-2 WHERE Winning team = a.z.k./roc-compétition AND Circuit = zolder |
| SQL (P) | SELECT (Date) FROM 2-17997366-2 WHERE Circuit = zolder AND Winning team = a.z.k./roc-compétition |
| ANS (T) | 5 may |
| ANS (P) | None |

**ERROR Question**

**69 2925**

What was the first Round with a Pick # greater than 1 and 140 Overall?

| TBL | “Round”, “Pick #”, “Overall”, “Name”, “Position”, “College” |
| SQL (T) | SELECT min(Round) FROM 2-15198842-23 WHERE Pick # > 1 AND Overall > 140 |
| SQL (P) | SELECT min(Round) FROM 2-15198842-23 WHERE Pick # > 1 AND Overall = 140 |
| ANS (T) | None |
| ANS (P) | 6.0 |
70 6224 NL On January 29, who had the decision of Mason?

TBL “Date”, “Visitor”, “Score”, “Home”, “Decision”, “Attendance”, “Record”

SQL (T) SELECT (Visitor) FROM 2-11756731-6 WHERE Decision = mason AND Date = january 29

SQL (P) SELECT (Decision) FROM 2-11756731-6 WHERE Date = january 29 AND Decision = mason

ANS (T) nashville

ANS (P) mason

71 4229 NL What venue had an event on 17 November 1963?

TBL “Season”, “Date”, “Winner”, “Score [C]”, “Venue”, “Competition round”

SQL (T) SELECT (Venue) FROM 2-17299309-4 WHERE Season = 1963 AND Date = 17 november 1963

SQL (P) SELECT (Venue) FROM 2-17299309-4 WHERE Date = 17 november 1963

ANS (T) estadio nacional

ANS (P) estadio nacional

72 7954 NL What was the attendance when the record was 77-54?

TBL “Date”, “Opponent”, “Score”, “Loss”, “Attendance”, “Record”

SQL (T) SELECT min(Attendance) FROM 2-12207430-6 WHERE Record = 77-54

SQL (P) SELECT (Attendance) FROM 2-12207430-6 WHERE Record = 77-54

ANS (T) 30,224

ANS (P) 30,224

73 3728 NL How many 2007’s have a 2000 greater than 56,6, 23,2 as 2006, and a 1998 greater than 61,1?


SQL (T) SELECT sum(2007) FROM 2-15348345-1 WHERE 2000 > 56,6 AND 2006 = 23,2 AND 1998 > 61,1


ANS (T) None

ANS (P) 0

74 625 NL What is the sexual abuse rate where the conflict is the Burundi Civil War?


SQL (T) SELECT min(Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = Burundi Civil War

SQL (P) SELECT (Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = burundi civil war

ANS (T) 80.0

ANS (P) 80.0

75 7290 NL What is the total poverty (2009) HPI-1 % when the extreme poverty (2011) < 1.25 US$ % of 16.9, and the human development (2012) HDI is less than 0.581?
SQL (P) SELECT sum(Rank) FROM 2-1615758-2 WHERE Year < 2002 AND Chassis = reynard 2ki
ANS (T) 19th
ANS (P) 19.0
ERROR None

What Nominating festival was party of the adjustment film?

TBL “Category”, “Film”, “Director(s)”, “Country”, “Nominating Festival”
SQL (T) SELECT (Nominating Festival) FROM 2-12152327-6 WHERE Film = adjustment
SQL (P) SELECT (Nominating Festival) FROM 2-12152327-6 WHERE Film = party of the adjustment
ANS (T) prix uip angers
ANS (P) None
ERROR None

What is the train number when the time is 10:38?

SQL (T) SELECT max(Train number) FROM 1-29202276-2 WHERE Time = 10:38
SQL (P) SELECT (Train number) FROM 1-29202276-2 WHERE Time = 10:38
ANS (T) 16381.0
ANS (P) 16381.0
ERROR Ground Truth

How many lines have the segment description of red line mos-2 west?

TBL “Segment description”, “Date opened”, “Line(s)”, “Endpoints”, “# of new stations”, “Length (miles)”
SQL (T) SELECT (Line(s)) FROM 1-1817879-2 WHERE Segment description = Red Line MOS-2 West
SQL (P) SELECT count(Line(s)) FROM 1-1817879-2 WHERE Segment description = red line mos-2 west
ANS (T) red, purple 1
ANS (P) 1
ERROR Ground Truth

In the match where the home team scored 14.20 (104), how many attendees were in the crowd?

TBL “Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”
SQL (T) SELECT sum(Crowd) FROM 2-10790397-5 WHERE Home team score = 14.20 (104)
SQL (P) SELECT (Crowd) FROM 2-10790397-5 WHERE Home team score = 14.20 (104)
ANS (T) 25.0
ANS (P) 25,000
ERROR Ground Truth

What is the grid number with less than 52 laps and a Time/Retired of collision, and a Constructor of arrows - supertec?

SQL (T) SELECT count(Grid) FROM 2-1123405-2 WHERE Laps < 52 AND Time/Retired = collision AND Constructor = arrows - supertec
SQL (P)  SELECT (Grid) FROM 2-1123405-2 WHERE Constructor = arrows AND Laps < 52 AND Time/Retired = collision

ANS (T)  1
ANS (P)  None
ERROR  Ground Truth

86 5893 NL  Name the team for launceston

TBL  “Race Title”, “Circuit”, “City / State”, “Date”, “Winner”, “Team”

SQL (T)  SELECT (Team) FROM 2-11880375-2 WHERE Race Title = launceston

SQL (P)  SELECT (Team) FROM 2-11880375-2 WHERE City / State = launceston

ANS (T)  shell ultra-hi racing
ANS (P)  None
ERROR  Ground Truth

87 627 NL  What is the sexual abuse rate where the conflict is the Second Sudanese Civil War?


SQL (T)  SELECT min(Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = Second Sudanese Civil War

SQL (P)  SELECT (Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = second sudanese civil war

ANS (T)  400.0
ANS (P)  400.0
ERROR  Ground Truth

88 7578 NL  What is the high checkout when Legs Won is smaller than 9, a 180s of 1, and a 3-dart Average larger than 88.36?

TBL  “Player”, “Played”, “Legs Won”, “Legs Lost”, “100+”, “140+”, “180s”, “High Checkout”, “3-dart Average”

SQL (T)  SELECT sum(High Checkout) FROM 2-18621456-22 WHERE Legs Won < 9 AND 180s = 1 AND 3-dart Average > 88.36

SQL (P)  SELECT max(High Checkout) FROM 2-18621456-22 WHERE Legs Won < 9 AND 180s = 1 AND 3-dart Average > 88.36

ANS (T)  None
ANS (P)  None
ERROR  Ground Truth

89 1056 NL  When the total score is 740, what is tromso?


SQL (T)  SELECT min(Tromsø) FROM 1-19439864-2 WHERE Total = 740

SQL (P)  SELECT (Tromsø) FROM 1-19439864-2 WHERE Total = 740

ANS (T)  70.0
ANS (P)  70.0
ERROR  Ground Truth

90 6089 NL  What are the draws when wins are fewer than 9 and byes fewer than 2?


SQL (T)  SELECT count(Draws) FROM 2-11338646-3 WHERE Wins < 9 AND Byes < 2

SQL (P)  SELECT avg(Draws) FROM 2-11338646-3 WHERE Wins < 9 AND Byes < 2

ANS (T)  0
ANS (P)  None
91 597
NL Name the year for sammo hung for ip man 2
TBL “Year”, “Best Film”, “Best Director”, “Best Actor”, “Best Actress”, “Best Supporting Actor”, “Best Supporting Actress”
SQL (T) SELECT (Year) FROM 1-15301258-1 WHERE Best Supporting Actor = Sammo Hung for Ip Man 2
SQL (P) SELECT (Year) FROM 1-15301258-1 WHERE Best Actor = sammo hung
ANS (T) 2011 5th
ANS (P) None

92 1122
NL What was the res for the game against Payam?
TBL “Date”, “Team #1”, “Res.”, “Team #2”, “Competition”, “Attendance”, “Remarks”
SQL (T) SELECT (Res.) FROM 1-2015453-1 WHERE Team #2 = Payam
SQL (P) SELECT (Res.) FROM 1-2015453-1 WHERE Team #1 = payam
ANS (T) 1–1
ANS (P) None

93 55
NL What are the races that johnny rutherford has won?
SQL (T) SELECT (Name) FROM 1-10706879-3 WHERE Winning driver = Johnny Rutherford
SQL (P) SELECT (Rd) FROM 1-10706879-3 WHERE Winning driver = johnny rutherford
ANS (T) kraco car stereo 200
ANS (P) 1.0

94 5746
NL How many goals were scored on 21 Junio 2008?
TBL “Goal”, “Date”, “Venue”, “Result”, “Competition”
SQL (T) SELECT count(Goal) FROM 2-1192553-1 WHERE Date = 21 junio 2008
SQL (P) SELECT (Goal) FROM 2-1192553-1 WHERE Date = 21 junio 2008
ANS (T) 1
ANS (P) 13.0

95 5705
NL What is the grid for the Minardi Team USA with laps smaller than 90?
SQL (T) SELECT (Grid) FROM 2-10823048-3 WHERE Laps < 90 AND Team = minardi team usa
SQL (P) SELECT sum(Grid) FROM 2-10823048-3 WHERE Team = minardi team usa AND Laps < 90
ANS (T) 12.0
ANS (P) 12.0

96 3992
NL What is the average number of gold medals when the total was 1335 medals, with more than 469 bronzes and more than 14 silvers?
TBL “Rank”, “Gold”, “Silver”, “Bronze”, “Total”
SQL (T) SELECT avg(Gold) FROM 2-15428689-2 WHERE Silver > 14 AND Total = 1335 AND Bronze > 469
SQL (P) SELECT sum(Grid) FROM 2-10823048-3 WHERE Team = minardi team usa AND Laps < 90
ANS (T) 12.0
ANS (P) 12.0
<table>
<thead>
<tr>
<th>SQL (P)</th>
<th>SELECT avg(Gold) FROM 2-15428689-2 WHERE Silver &gt; 14 AND Bronze &gt; 469 AND Total = 1335 medals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANS (T)</td>
<td>None</td>
</tr>
<tr>
<td>ANS (P)</td>
<td>None</td>
</tr>
<tr>
<td>ERROR</td>
<td>None</td>
</tr>
</tbody>
</table>

97 6729 NL What was the year that had Anugerah Bintang Popular Berita Harian 23 as competition?

TBL “Year”, “Competition”, “Awards”, “Category”, “Result”

SQL (T) SELECT count(Year) FROM 2-17838670-5 WHERE Competition = anugerah bintang popular berita harian 23

SQL (P) SELECT (Year) FROM 2-17838670-5 WHERE Competition = anugerah bintang popular berita harian 23

ANS (T) 1

ANS (P) 2010.0

ERROR Ground Truth

98 7479 NL What’s the position that has a total less than 66.5m, a compulsory of 30.9 and voluntary less than 33.7?

TBL “Position”, “Athlete”, “Compulsory”, “Voluntary”, “Total”

SQL (T) SELECT min(Position) FROM 2-18662083-1 WHERE Total < 66.5 AND Compulsory = 30.9 AND Voluntary < 33.7

SQL (P) SELECT sum(Position) FROM 2-18662083-1 WHERE Compulsory = 30.9 AND Voluntary < 33.7 AND Total < 66.5

ANS (T) None

ANS (P) None

ERROR Ground Truth

99 1263 NL What episode had 10.14 million viewers (U.S.)?

TBL “No.”, “#”, “Title”, “Directed by”, “Written by”, “U.S. viewers (million)”, “Original air date”, “Production code”

SQL (T) SELECT min(#) FROM 1-21550897-1 WHERE U.S. viewers (million) = 10.14

SQL (P) SELECT (Title) FROM 1-21550897-1 WHERE U.S. viewers (million) = 10.14

ANS (T) 11.0

ANS (P) “arrow of time”

ERROR Ground Truth

100 557 NL Name the english gloss for ha´ána

TBL “English gloss”, “Santee-Sisseton”, “Yankton-Yanktonai”, “Northern Lakota”, “Southern Lakota”

SQL (T) SELECT (English gloss) FROM 1-1499774-5 WHERE Santee-Sisseton = ha´ána

SQL (P) SELECT (English gloss) FROM 1-1499774-5 WHERE Southern Lakota = ha´ána

ANS (T) morning

ANS (P) None

ERROR Question